

BERETTA STEFANO

ste.beretta@gmail.com ◇ 340 0823140

via Papa Giovanni XXIII 31/A, 20842, Besana in Brianza (MB), Italy

TITOLI DI STUDIO

Dottorato di Ricerca in Informatica (XXV ciclo)

01/2010 - 02/2013

Università degli Studi di Milano - Bicocca, Dipartimento di Informatica, Sistemistica e Comunicazione (DISCo)

- Il titolo della tesi è “*Algorithms for Next Generation Sequencing Data Analysis*”, svolta sotto la supervisione di Prof.ssa Paola Bonizzoni e Prof. Gianluca Della Vedova.

Laurea Specialistica in Informatica

10/2006 - 04/2009

Università degli Studi di Milano - Bicocca, Dipartimento di Informatica, Sistemistica e Comunicazione (DISCo)

- Il titolo della tesi è “*Algoritmi per la Predizione della Struttura Secondaria dell'RNA*”, svolta sotto la supervisione di Prof.ssa Paola Bonizzoni e Prof. Gianluca Della Vedova.
- Voto: 110/110 e Lode

Laurea Triennale in Informatica

09/2003 - 10/2006

Università degli Studi di Milano - Bicocca, Dipartimento di Informatica, Sistemistica e Comunicazione (DISCo)

- Il titolo della relazione finale è “*Un framework per la simulazione di sistemi complessi: metodologie di visualizzazione scientifica*”, svolta sotto la supervisione di Prof. Gianpiero Cattaneo e Dott. Fabio Mario Farina.
- Voto: 107/110

Diploma Magistrale

09/1998 - 07/2003

Istituto tecnico commerciale sperimentale - Alessandro Greppi, via Dei Mille 31, Monticello B.za (LC).

- Informatica, Elettronica, Sistemistica.

POSIZIONI

Assegno di Ricerca

05/2013 - 05/2014

Consiglio Nazionale delle Ricerche (CNR), Istituto di Tecnologie Biomediche

- Ricerca nell'ambito del progetto *MERIT* "MEDical Research in ITaly."

Collaborazione Coordinata e Continuativa

05/2014 - 05/2015

Consiglio Nazionale delle Ricerche (CNR), Istituto di Tecnologie Biomediche

- Collaborazione di Ricerca nell'ambito del progetto *FIRB HIRMA*.

ATTIVITÀ SCIENTIFICA

Descrizione dell'attività di ricerca

- La mia attività di ricerca si colloca in ambito bioninformatico. I miei principali interessi riguardano il disegno, lo sviluppo e la sperimentazione di algoritmi combinatori per la soluzione di problemi in genomica e trascrittomica. Nello specifico, lavoro alla realizzazione di applicazioni per l'analisi di sequenza biologiche (DNA ed RNA).

Durante il mio dottorato di ricerca mi sono occupato dello sviluppo di un modello computazionale per la caratterizzazione di eventi di splicing alternativo, a partire da dati di sequenziamento di nuova generazione (RNA-Seq). Obiettivo di tale ricerca è quello di ricostruire a partire da dati RNA-Seq, in presenza o assenza di un genoma di riferimento, la struttura dei geni espressi. Tale struttura è finalizzata a rappresentare la composizione di trascritti dei geni, che sono soggetti ai meccanismi di splicing alternativo.

Ho inoltre svolto ricerca su pipeline per l'analisi di dati di sequenziamento trascrittomici, combinando metodi per l'allineamento di sequenze come TopHat e GSNAP, a tool per la ricostruzione di trascritti, come ad esempio Cufflinks e SLIDE. Tali pipeline sono state valutate mediante una comparazione diretta con dati simulati e reali, provenienti dal progetto ENCODE, sulla base dei risultati ottenuti.

Attualmente la ricerca è orientata al perfezionamento del tool sviluppato nell'attività di tesi ed in futuro si vuole approfondire la ricostruzione di trascritti facendo riferimento ad annotazioni di cDNA o basandosi su informazioni provenienti da altri genomi di riferimento.

Durante l'attività di ricerca svolta ho acquisito competenze relative all'utilizzo di banche dati genomiche, come ad esempio NCBI, sia in termini di interrogazione e navigazione finalizzata al reperimento dei dati che di manipolazione di questi ultimi.

Sono anche interessato all'integrazione di tool di predizione della struttura genica con tool di predizione di eventi di variazione strutturale, con l'obiettivo di integrare i due approcci al fine di migliorare la qualità dei risultati prodotti.

Infine, durante il mio periodo presso la *Technical University of Catalonia* (UPC) sotto la supervisione di Prof. Gabriel Valiente, mi sono occupato di studi relativi a problemi di metagenomica. In particolare ho lavorato allo sviluppo di algoritmi per il corretto assegnamento di sequenze genomiche, provenienti da organismi presenti in un campione analizzato, in alberi tassonomici. Questo lavoro di carattere sia algoritmico che applicativo, mi ha permesso di sviluppare delle competenze in ambito metagenomico riguardanti tecniche di assegnamento di sequenze e comparazione di alberi tassonomici. A questo proposito, mi sono occupato dell'analisi delle principali banche dati metagenomiche, quali NCBI Taxonomy, RDP, Greengenes e SILVA.

Dopo la conclusione del mio dottorato di ricerca ho trascorso un anno come assegnista di ricerca al Consiglio Nazionale delle Ricerche (CNR) presso la sede dell'Istituto di Tecnologie Biomediche (ITB). Durante questo periodo sono stato coinvolto in diversi progetti che prevedevano la realizzazione di tecniche bioinformatiche orientate all'analisi di dati di sequenziamento in ambito clinico.

In questo modo ho potuto migliorare notevolmente le mie conoscenze rispetto agli aspetti applicativi ed inoltre sono venuto a contatto con diverse realtà bioinformatiche che si occupano di ricerca in ambito clinico. Queste collaborazioni mi hanno permesso di comprendere meglio le problematiche legate all'impiego di informazioni ottenute con sequenziamento di nuova generazione e di estendere la mia conoscenza rispetto a questa tipologia di dati.

ELENCO DELLE PUBBLICAZIONI

Articoli su Riviste Internazionali

- [1] **Stefano Beretta**, Paola Bonizzoni, Gianluca Della Vedova, Yuri Pirola, and Raffaella Rizzi. "Modeling Alternative Splicing Variants from RNA-Seq Data with Isoform Graphs". *Journal of Computational Biology* 21.1 (2014), 16–40. DOI: 10.1089/cmb.2013.0112.
- [2] Mauro Castelli, **Stefano Beretta**, and Leonardo Vanneschi. "A hybrid genetic algorithm for the repetition free longest common subsequence problem". *Operations Research Letters* 41.6 (2013), 644–649. ISSN: 0167-6377. DOI: 10.1016/j.orl.2013.09.002.

- [3] Daniel Alonso-Aleman, Aurélien Barré, **Stefano Beretta**, Paola Bonizzoni, Macha Nikolski, and Gabriel Valiente. “Further Steps in TANGO: Improved Taxonomic Assignment in Metagenomics”. *Bioinformatics* 30.1 (2014), 17–23. DOI: 10.1093/bioinformatics/btt256.

Atti di Conferenze Internazionali

- [4] **Stefano Beretta**, Paola Bonizzoni, Raffaella Rizzi, and Gianluca Della Vedova. “Reconstructing isoform graphs from RNA-Seq data”. In: *Bioinformatics and Biomedicine (BIBM), 2012 IEEE International Conference on*. Oct. 2012, 1–4. DOI: 10.1109/BIBM.2012.6392734.
- [5] Stefano Beretta and Riccardo Dondi. “Gene Tree Correction by Leaf Removal and Modification: Tractability and Approximability”. English. In: *Language, Life, Limits*. Ed. by Arnold Beckmann, Erzsébet Csuhaj-Varjú, and Klaus Meer. Vol. 8493. Lecture Notes in Computer Science. Springer International Publishing, 2014, 42–52. ISBN: 978-3-319-08018-5. DOI: 10.1007/978-3-319-08018-5_2.
- [6] Niko Beerenwinkel, **Stefano Beretta**, Paola Bonizzoni, Riccardo Dondi, and Yuri Pirola. “Covering Pairs in Directed Acyclic Graphs”. In: *LATA*. Ed. by Adrian Horia Dediu, Carlos Martín-Vide, José Luis Sierra-Rodríguez, and Bianca Truthe. Vol. 8370. Lecture Notes in Computer Science. Springer, Mar. 2014, 126–137. ISBN: 978-3-319-04920-5. DOI: 10.1007/978-3-319-04921-2_10.

Workshop Internazionali (con peer-review)

- [7] Anna Paola Carrieri, **Stefano Beretta**, Gianluca Della Vedova, Ernesto Picardi, Yuri Pirola, Raffaella Rizzi, Graziano Pesole, and Paola Bonizzoni. “An in-silico framework for comparing and validating transcripts predicted from single and paired-end reads”. In: *Next Generation Sequencing Workshop*. Oct. 2012.
- [8] Yuri Pirola, Raffaella Rizzi, **Stefano Beretta**, Ernesto Picardi, Graziano Pesole, Gianluca Della Vedova, and Paola Bonizzoni. “PItronNext: a fast method for detecting the gene structure due to alternative splicing via ESTs, mRNAs, and RNA-Seq data”. In: *EURASNET Symposium on Regulation of Gene Expression through RNA Splicing*. Mar. 2012.
- [9] **Stefano Beretta**, Paola Bonizzoni, Raffaella Rizzi, and Gianluca Della Vedova. “Identification of Alternative Splicing variants from RNA-seq Data”. In: *Next Generation Sequencing Workshop*. Oct. 2011.
- [10] **Stefano Beretta**, Paola Bonizzoni, Raffaella Rizzi, and Gianluca Della Vedova. “Alternative Splicing from RNA-seq Data without the Genome”. In: *ISMB/ECCB - 8th Special Interest Group meeting on Alternative Splicing (AS-SIG)*. July 2011.

Tesi di Dottorato di Ricerca

- [11] **Stefano Beretta**. “Algorithms for next generation sequencing data analysis”. PhD thesis. Università degli Studi di Milano-Bicocca, Feb. 2013.

COLLABORAZIONI

Partecipazione a Progetto di Ricerca

- PRIN 2010 - 2011 dal titolo “*Automi e Linguaggi Formali: Aspetti Matematici e Applicativi*”. Coordinatore scientifico: Prof. Antonio Restivo. Unità: Università degli Studi di Milano - Bicocca, responsabile scientifico: Prof.ssa Paola Bonizzoni.
- FIRB HIRMA codice RBAP11YS7K
- MERIT - “MEdical Research in ITaly.”

Borsa di Studio

- Borsa di Studio Ministeriale per la frequenza del Dottorato di Ricerca da Gennaio 2010 a Dicembre 2012.

EVENTI E CONVEGNI

Partecipazione a Comitati Organizzativi

- Comitato organizzativo della conferenza *Computability in Europe (CiE) - The Nature of Computation*, 1-5 Luglio 2013, Università degli Studi di Milano-Bicocca.

Relazione a Convegni Internazionali

- Partecipazione e intervento a *Bioinformatics and Biomedicine (BIBM)*, 2012 IEEE International Conference, 4-7 Ottobre 2012, Philadelphia PA, USA, in qualità di relatore del lavoro [4].
- Partecipazione e intervento a *Next Generation Sequencing Workshop*, 12-14 Ottobre 2011, Bari, Italia, in qualità di relatore del lavoro [9].
- Partecipazione e intervento a *ISMB/ECCB - 8th Special Interest Group meeting on Alternative Splicing (AS-SIG)*, 15-16 Luglio 2011, Vienna, Austria, in qualità di relatore del lavoro [10].

Partecipazione a Convegni Internazionali

- Partecipazione a *RECOMB - Conference on Research in Computational Molecular Biology*, Aprile 2012, Barcellona, Spagna.

Scuole

- Frequenza della scuola estiva Lipari International Summer School on Bioinformatics and Computational Biology, dal titolo "*Biological Sequence Analysis and High Throughput Technologies*", organizzata dall'Università degli Studi di Catania, 2011.

Visite di Ricerca

- *Visiting Scholar* presso la Technical University of Catalonia (UPC) in Barcellona, Spagna, da Aprile 2012 a Luglio 2012, sotto la supervisione di Prof. Gabriel Valiente.

ATTIVITÀ DIDATTICHE

Docente di lezioni frontali

- Insegnamento corso "Informatica di Base 1" (60 ore), corso di Laurea in Lingue e Letterature Straniere, a.a. 2011/2012.
Università degli studi di Bergamo - facoltà di Lingue e Letterature Straniere. Piazza Rosate, 2 - 24129 Bergamo.
- Insegnamento corso "Informatica di Base 1" (60 ore), corso di Laurea in Lingue e Letterature Straniere, a.a. 2010/2011.
Università degli studi di Bergamo - facoltà di Lingue e Letterature Straniere. Piazza Rosate, 2 - 24129 Bergamo.

Esercitazioni

- Esercitazioni di laboratorio (24 ore), insegnamento "Laboratorio di Algoritmi e Strutture Dati", corso di Laurea in Informatica, a.a. 2010/2011.
Università degli studi di Milano - Bicocca. Piazza dell'Ateneo Nuovo, 1 - 20126 Milano.
- Esercitazioni di laboratorio (24 ore), insegnamento "Laboratorio di Informatica I", corso di Laurea in Fisica, a.a. 2010/2011.
Università degli studi di Milano - Bicocca. Piazza dell'Ateneo Nuovo, 1 - 20126 Milano.

- Esercitazioni di laboratorio (24 ore), insegnamento “Laboratorio di Matematica e Informatica I”, corso di Laurea in Matematica, a.a. 2010/2011.
Università degli studi di Milano - Bicocca. Piazza dell’Ateneo Nuovo, 1 - 20126 Milano.
- Esercitazioni di recupero per il corso “Algoritmi e Ricerca Operativa” (12 ore), corso di Laurea in Informatica, a.a. 2009/2010.
Università degli studi di Milano - Bicocca. Piazza dell’Ateneo Nuovo, 1 - 20126 Milano.
- Esercitazioni di laboratorio per l’insegnamento “Laboratorio di Algoritmi e Strutture Data” (24 ore), corso di Laurea in Informatica, a.a. 2009/2010.
Università degli studi di Milano - Bicocca. Piazza dell’Ateneo Nuovo, 1 - 20126 Milano.

Altre esperienze didattiche di livello universitario

- Tutor all’interno del Progetto Sovvenzione Globale Learning Week Id 158 - titolo: *Informatica ieri, oggi e domani* - edizione n. 286, 08/02/2010 - 12/02/2010.
C.I.L.F (Centro per l’innovazione del Lavoro e della Formazione) Società Cooperativa Sociale. via Paolo Bassi, 3 - Milano.

Supervisione Stage

- Co-supervisione di Stage, effettuati a conclusione del Corso di Laurea in Informatica, presso l’Università degli Studi di Milano-Bicocca.

ESPERIENZA PROFESSIONALE

Prestazione di Lavoro Occasionale

05/2008 - 05/2009

Bytech - servizi informatici, via Cesare Cantù 83, Cinisello Balsamo (MI)

- Supporto Help desk in Bio-Rad Laboratories IT Operations per sedi Europee (Italia, Spagna, Portogallo e Grecia), presso la sede di Segrate (MI)

Attività di tutorato presso i Laboratori Didattici d’Ateneo

10/2007 - 03/2008

Università degli studi di Milano - Bicocca. Piazza dell’Ateneo Nuovo, 1 - 20126 Milano

- Supporto informatico all’utenza del laboratorio: supporto all’utilizzo degli applicativi e risoluzione di problemi software

Attività di tutorato presso i Laboratori Didattici d’Ateneo

04/2007 - 09/2007

Università degli studi di Milano - Bicocca. Piazza dell’Ateneo Nuovo, 1 - 20126 Milano

- Supporto informatico all’utenza del laboratorio: supporto all’utilizzo degli applicativi e risoluzione di problemi software

CAPACITÀ E COMPETENZE PROFESSIONALI

Settore di competenza

- Algoritmi e Bioinformatica. I miei interessi riguardano principalmente il disegno e lo sviluppo di algoritmi di carattere combinatorio, con particolare attenzione ai problemi in ambito bioinformatico. Mi occupo inoltre dello studio e della realizzazione di tool applicativi per l’analisi di sequenze per problematiche biologiche. Mi interessa anche a studi di complessità computazionale.

Capacità e competenze sociali

- Buone capacità relazionali acquisite attraverso la partecipazione a bandi 150 ore di tutoraggio, presso l’Università degli Studi di Milano - Bicocca. Capacità di lavorare in gruppo, anche coordinata a distanza tramite Internet, gestendo tempo e risorse. Disponibilità al confronto ed alla risoluzione condivisa di problemi.

Capacità e competenze organizzative

- Buona attitudine a collaborare con altri nella gestione e nello sviluppo di progetti, rispettando i tempi di consegna, ottenuta grazie al lavoro in team in ambito universitario. Sono in grado di organizzare autonomamente il lavoro definendo le priorità e assumendo responsabilità. Ho coordinato e coordino le attività di diversi studenti (tesi e altri progetti).

Capacità e competenze tecniche

- Ottime propensioni scientifiche e matematiche. Capacità di problem solving ed impiego di tecniche algoritmiche.

Capacità e competenze informatiche

- Durante il mio dottorato di ricerca in informatica ho sviluppato competenze informatiche in diversi linguaggi programmatici e di scripting, quali C/C++, Java, Perl e Python. Ho inoltre acquisito conoscenze dei linguaggi R e Gnuplot per l'analisi statistica e del linguaggio TeX/LaTeX per la realizzazione di documenti testuali.

Ho inoltre acquisito le seguenti competenze:

- Ottima conoscenza dei sistemi operativi Windows/Linux.
- Ottima conoscenza dei pacchetti MS Office e OpenOffice.
- Ottima conoscenza dei seguenti ambienti e tool di sviluppo: Eclipse, EMACS, GCC.
- Ottima conoscenza di strumenti di versioning GIT, CVS e SVN.
- Buona conoscenza delle principali piattaforme e-learning: Moodle e IBM Quickr.
- Conoscenza di strumenti di generazione di documentazione Doxygen.
- Conoscenza base dei protocolli di rete TCP/IP.

Luogo e Data

Milano, 19/09/2014

FIRMA
